Statistical Inference Course Project Part 1

Assignment Description

Investigate the exponential distribution in R and compare it with the Central Limit Theorem.The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

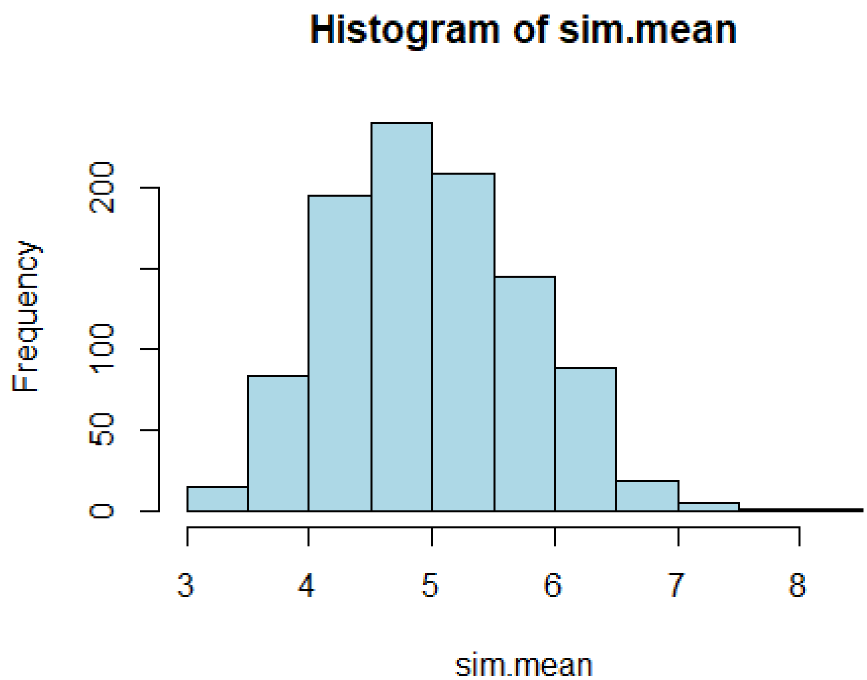
Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should:

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Simulation Excercise

1.Show the sample mean and compare it to the theoretical mean of the distribution.

```
set.seed(1)
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.5.2

sim.n<-1000
lambda<-0.2
n<-40
sim.matrix<-matrix(rexp(sim.n*n,rate = lambda),sim.n,40)

sim.mean<-apply(sim.matrix,1,mean)

hist(sim.mean,col="lightblue")
```




```
sample.mn<-mean(sim.mean)
sample.sd<-sd(sim.mean)

theory.mn<-1/lambda
theory.sd<-1/lambda

sample.mn;theory.mn

## [1] 4.990025

## [1] 5
```

Center of the ditribution made by simulation is 4.990025 whereas theoretical mean is 5 with lambda=0.2. It implies that CLT works well and it generates pretty close estimation.

2.Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
sample.var<-var(sim.mean)
theory.var<-(1/lambda)^2/n

sample.var;theory.var

## [1] 0.6177072

## [1] 0.625
```

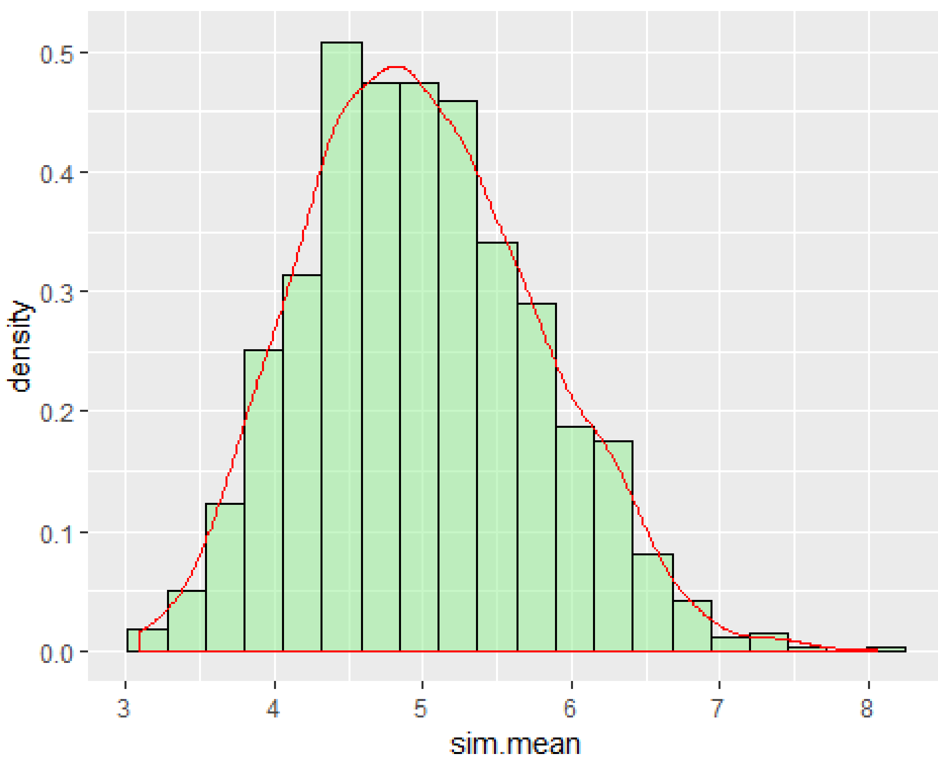
Variance based on simulation is 0.6177072 while theoretical variance is 0.625. Both are close enugh.

3.Show that the distribution is approximately normal.

- Density and Histogram
- QQplot

Histogram and Density

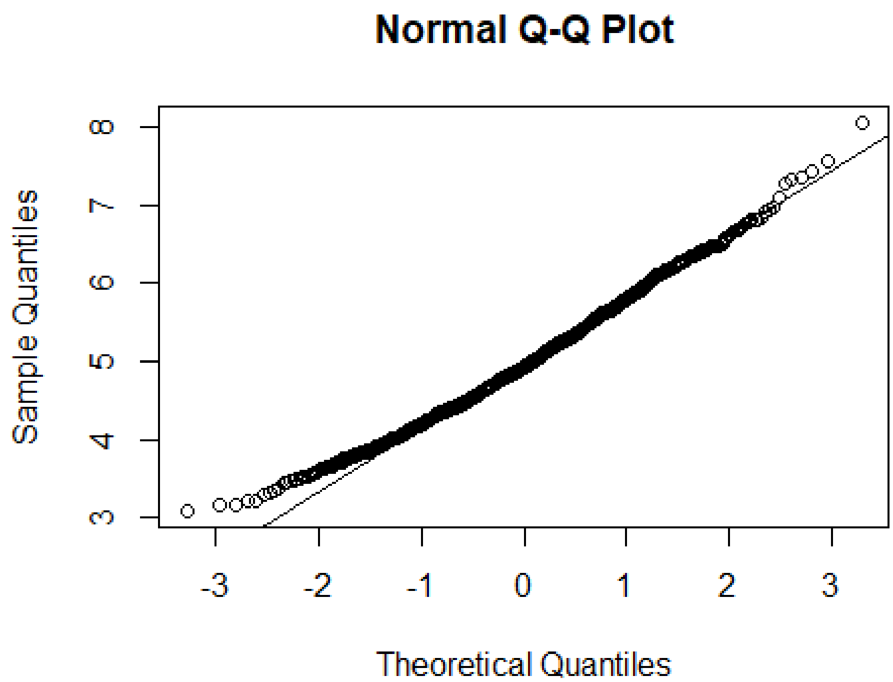
```
sim.mean<-as.data.frame(sim.mean)
sim.mean.n<-as.numeric(unlist(sim.mean))

ggplot(sim.mean,aes(x=sim.mean))+

geom_histogram(aes(y=..density..),fill="lightgreen",col="black",alpha=0.5,pos
ition = "dodge",bins = 20)+
    geom_density(aes(y=..density..),col='red',alpha=0.5)
```



The distribution of the mean of simulation looks very normally distributed with the mean of 4.9 and the variance of 0.6

QQPlot

```
qqnorm(sim.mean$sim.mean);qqline(sim.mean$sim.mean)
```



Based on the plotted graph, it seems pretty normally distributed since, data points are reasonably closely scattered aruond the qqline.